

## **TRI DATA COLLECTION, PROCESSING and MANAGEMENT** **for Stakeholder Outreach**

### **I. BACKGROUND**

This is one of three papers which describe aspects of the Toxics Release Inventory (TRI) Program and raise issues for stakeholder input. The scope of each paper corresponds to a phase of the annual TRI reporting cycle. TRI data for a calendar year must be reported to EPA each year by July 1<sup>st</sup> after the end of the year. Therefore, reporting years are the same as calendar years. The “reporting cycle” begins with EPA’s compliance assistance activities, including the development of its reporting forms and instructions package that is mailed to facilities generally in March of each year. Once EPA via the EPCRA Reporting Center (managed by a contractor for the Agency) receives the TRI submissions (either by a paper form, diskette, or via the internet), the data are entered (over 91,000 forms in 2000) in the TRI database known as TRIS (Toxics Release Inventory System). After entry into the database, EPA runs a number of data quality checks on both the facility identification information and on the chemical-specific data. After the data entry and data quality steps are completed, the TRI database is “frozen” for analysis and development of data products for release to the public. Generally, EPA announces the annual release of the TRI data by holding a press event or issuing a press release, and simultaneously notifying a wide range of stakeholders.

This first background paper for this stakeholder process is entitled *TRI Data Collection, Processing and Management*, and addresses the TRI data process beginning with the receipt of TRI submissions received by the EPCRA Reporting Center and ending at the data “freeze.” The second paper, *TRI Data Release Issue Paper*; discusses TRI data products, the process for analyzing and releasing the TRI data, uses of the data, and issues and considerations associated with these aspects of the TRI program. The third paper is *TRI Compliance Assistance Activities*. TRI compliance assistance activities are carried out throughout the year with certain of the activities being closely aligned with the reporting cycle.

The purpose of this issue paper is to describe the operations of the EPCRA Reporting Center and to provide insight on how a TRI submission received via paper, disk or CDX, is processed. In each section we have included questions in italics to help focus the stakeholder comments.

The processing of the TRI data can be divided into seven steps:

1. Mailroom/records Management
2. Data capture, verification, validation and reconciliation
3. Data validation and Facility Data Profile (i.e., echo-back) preparation

4. Preparation, tracking and mailouts of Notices of Significant Errors (NOSE) and Notices of Noncompliance (NON)
5. Data Quality Exercise: State Reconciliation Reports
6. Data “Chill” and Data Quality Steps
7. Data “Freeze”

Please see the flow chart at the end of this document for a simplified process of the data flow for a TRI submission whether received via paper, disk or CDX. More details on each of the steps required to process TRI submissions follow.

## **II. OVERVIEW OF THE EPCRA REPORTING CENTER**

The EPCRA Reporting Center is responsible for managing the full life cycle of submissions received for the TRI Program. This includes all forms, documents, electronic submissions, and any other communications received at the EPCRA Reporting Center (e.g., paper Form R and Form A certification statement; diskettes, letters regarding a facility change of address, etc.). These total approximately 120,000 items annually (per calendar year), representing submissions and revisions from all reporting years. The EPCRA Reporting Center is responsible for the processing, tracking, data entry, data quality and reconciliation checks of all data received.

Submissions may be hard copy Form Rs and Form A certification statements, diskette submissions, and submissions received via the internet through the Agency’s Central Data Exchange (CDX). Also received are responses to Facility Data Profiles (i.e., echo-back), to NOTES and NOSEs, letters and miscellaneous correspondence, and requests for withdrawals and revisions. Certification letters are generally included with the diskette submission. For internet submissions, these are received separately.

About 80% to 90% of original submissions are received between July 1 and July 31, which is considered “peak season” for the EPCRA Reporting Center. Another increase in mail is experienced in the Fall timeframe when Facility Data Profile responses are returned for processing. However, the receipt and processing of TRI data is a year-round process, as revisions, facility changes, and other correspondence are received by the EPCRA Reporting Center throughout the year.

## **III. DATA PROCESSING STEPS**

### **Step 1: Mailroom/Records Management**

This step is intended to capture processing of the mail once it is received at the EPCRA

Reporting Center Mailroom.

The EPCRA Reporting Center Mailroom, also known as the Document Receipt and Identification Department, serves as the center of activity. All submissions and mail received for the TRI Program sent to the EPCRA Reporting Center go to the mailroom as a first step.

There are several categories of mail that the EPCRA Reporting Center receives. These categories are:

- Trade Secret Submissions (20 forms/RY00)
- Paper TRI Submissions (23,116 forms for RY00)
- Magnetic Media (disk) Submissions (80,133 forms for RY00)
- Certification Letters
- Miscellaneous Documents including:
  - Withdrawal requests
  - Freedom of Information Act (FOIA) requests
  - Facility Data Profiles Responses (28,960 FDPs generated, 6,529 responses)
  - General correspondence which includes: letters informing EPA of a change in facility name, facility address, facility closure, etc.. (1,396 documents)

Once all of the mail/items are opened, coded, and date stamped, they are entered into the Records Management System (RMS). All envelopes and their contents are date stamped, so the EPCRA Reporting Center may accurately track the receipt of all submissions and other documentation. RMS stores information such as postmarked date, actual receipt date, and the type of mail carrier used. Once this information is entered in RMS, the submission leaves the mailroom and goes to the Data Entry area, where it is entered or loaded (via diskette or downloaded from CDX) into TRIS.

*EPA would like comments and suggestions that would enhance our current procedures. Do you have any comments on the current procedures for our mailroom and records management? Comments can relate to the process as a whole or specific aspects of the mailroom or records management.*

## **Step 2: Data capture, verification, validation and reconciliation**

This step is intended to capture, verify, validate, and reconcile data the Reporting Center has received by magnetic media or paper.

TRI submissions can be received via paper, diskette, or by CDX. A paper Form R or Form A certification statement involves manual keying of the data submitted on paper forms into TRIS. For reporting year (RY) 2000, approximately 20% of submissions received were via paper. It takes approximately 30 minutes to process a paper submission. The EPCRA Reporting Center also received TRI submissions on diskette from either ATRS, TRI-ME, or commercially available vendor software. The processing time for a disk submission is approximately 8 minutes. A diskette may contain information on one or multiple facilities, multiple chemicals, and/or multiple reporting years for one facility. For RY 2000, the EPCRA Reporting Center received over 17,000 diskettes. Before a disk can be loaded in TRIS, the following steps occur:

- All disks are scanned for computer viruses and disk infected with a virus are removed.
- Any unprocessable (bad) disks are identified (e.g., disk cannot be read, is damaged in the mail, contains something other than a TRI submission, has a virus and the disk cannot be cleaned).
- The facility is notified via a “Bad Disk” letter.

In addition, TRI submissions can be received on the internet via the Agency’s portal, the Central Data Exchange (CDX). A submission through CDX takes approximately 9 minutes. The EPCRA Reporting Center downloads, on a daily basis during the peak season, all the submissions received via CDX into TRIS. These submissions are automatically virus-checked upon downloading. Once transferred into TRIS, these submissions received via CDX are tracked like any other submission received via disk.

The first steps of data capture (for a paper submission) is the entering of “facility information.” This information is contained in Part I of a Form R and the Form A certification statement. This is known as “Facility Data Capture.” For submissions received via disk or CDX, all of the information (both at the facility and chemical level) are entered in TRIS at the same time. The next step for all submissions, whether received via paper, disk or CDX is “Facility Reconciliation”.

The purpose of Facility Reconciliation is to verify and manually correct (if necessary), submitted facility data that does not match earlier submissions from the same facility in TRIS and to ensure that submissions are attached to the appropriate facility. A simple example relates to a facility’s TRI facility identification number (e.g., TRIFID). In RY 2000, Joe’s Painting Facility reported and their TRIFID was 12345JPNTN123MA. When the RY 2001 TRI submissions was received, the same facility submitted their TRIFID as: 12344JPNTN123MA. Facility reconciliation catches the incorrect TRIFID and is corrected in this process.

For a paper submission, after facility reconciliation is complete, the chemical information contained in Part II of the Form R is entered. This is termed “Chemical Data Capture.” For all paper

submissions, the next step is known as “Data Verification”. This is a double key entry of all data for Part II of a Form R.

### **Step 3: Data validation and Facility Data Profile (i.e., echo-back) preparation**

This step is intended to correct errors that prevent the form from being entered into the database, identify errors that adversely affect the quality of the TRI data, and notify submitters of simple clerical errors that EPA has corrected.

After the data is entered into TRIS, a “data validation” or error logic check is run on all of the data. This logic checks for a number of possible errors in a submission. One simple example is the submission of incomplete information, such as the omission of the telephone number for the technical contact. These errors are noted and as described below, a notice of the error will be sent to the facility.

After the data validation check is complete, one final processing step is “duplicate check”. During “duplicate check”, the goal is to identify and eliminate potential duplicate submissions for a particular reporting year. Ideally, a facility should send in one submission per chemical for each reporting year. However, some facilities submit both a paper Form R and a Form R via disk for the same chemical. The goal of duplicate check is to eliminate these redundancies. After duplicate check is complete, the next step is the generation of the Facility Data Profile (FDP), or an echo-back of the information received by EPA.

The FDP serves two primary purposes. First, EPA wants to provide the reporting facility the opportunity to confirm that the EPCRA Reporting Center has entered its data correctly into TRIS. Thus, the EPCRA Reporting Center “echoes-back” the information it has received. Second, if the EPCRA Reporting Center identifies potential errors in the submissions through the data validation step, the FDP indicates what these errors are and requests that the facility provide EPA with corrections. If the data presented on the FDP do not match those on the form(s) submitted by the facility, or if the EPCRA Reporting Center has identified errors in a facility’s submission(s), the facility may use the FDP to make the needed corrections. During RY 2000, 28,960 notices were generated. Of the 28,960 generated, 6,529 received responses. Within a FDP notice, there may be up to three different types of errors identified. These are discussed below:

(1) A Non-Technical Data Change (NDC) notifies the facility of a simple, clerical error that EPA has corrected for the facility. The EPCRA Reporting Center will correct simple, clerical errors that are not technical or scientific. For example, if a facility transposes the CAS number for sodium nitrite (e.g., the facility submits 7623-00-0 instead of 7632-00-0), the EPCRA Reporting Center will correct this clerical error and display the correct information on the FDP. It is not necessary to respond to a

NDC. EPA produced 2,653 NDCs in TRIS validation during the processing of RY 2000 data.

(2) A Notice of Technical Error (NOTE) highlights inconsistencies or miscalculations that may distort a facility's information in EPA's public data products, or skew analyses. A missing or invalid SIC code, or the use of range codes to reports PBT chemical releases and other waste management quantities are all examples of a NOTE level error. NOTE level errors should be responded to as soon as possible. EPA produced 15,024 NOTES during the processing of RY 2000 data.

(3) A Notice of Significant Error (NOSE) identifies errors that prevent a facility's submission from being entered in TRIS, or because it is missing critical information such as a chemical identifier or submission certification. A facility must respond to a NOSE within 21 days of receipt. EPA produced 96 NOSEs during the processing of RY 2000 data.

After the generation of the FDP, the submission(s) is now considered complete and is ready to be loaded to Envirofacts and TRI Explorer. Prior to the Public Data Release (PDR), the EPCRA Reporting Center tries to include all responses to FDPs received prior to the "Freeze" of TRIS (Freeze is explained in Step 7). These changes are reflected in the PDR, and in Envirofacts and TRI Explorer.

*[Please note: EPA is currently in the process of developing an online FDP/Echoback process that will allow facilities to login on a designated website to receive their reports. This enhancement replaces hard copy mailing of FDPs done in prior years. NOSEs and NONs will still be mailed by conventional means.]*

*EPA would like comments on TRI's procedures to capture, verify, validate, and reconcile data (Steps 2 and 3). Suggestions and comments can relate to TRI expediting this process. Also, we would like comments on our current and proposed FDP process. Suggestions and comments should focus on the idea of an online FDP. Will an online FDP be beneficial for facilities to access their data?*

#### **Step 4: Preparation, tracking and mailouts of Notices of Significant Errors (NOSE) and Notices of Noncompliance (NON)**

This step is intended to notify facilities of NOSEs and NONs.

As mentioned in Step 3, FDPs may identify one of three possible different types of errors. If a facility receives a NOSE(s) in their FDP, they should respond to EPA within 21 days of receipt. If a facility fails to respond to the NOSE(s) within 21 days of receipt, a Notice of Noncompliance (NON) could result.

A NON will require a facility to take the corrective action noted in the NOSE via the FDP within the 21 days and respond to the Agency. If a facility fails to respond to the FDP within the required time period, the Agency may take further action. During RY 2000, 42 NONs were issued by the Reporting Center.

The EPCRA Reporting Center is responsible for the tracking and generating of all NOSEs and NON letters. However, EPA makes the final determination whether to issue a NOSE and NON. All NON letters are mailed by certified mail and are a joint effort with the Office of Enforcement Compliance and Assurance (OECA).

*EPA would like comments regarding our current procedures to prepare, track, and mail the NOTES and NONs.*

### **Step 5: Data Quality Exercise: State Reconciliation Reports**

This step is intended to provide states information received by the TRI Program. This gives States the opportunity to ensure their TRI data is consistent with those received by EPA.

One major data quality step, is the State Reconciliation Reports. EPA sends the State Reconciliation Report via email to every State. This report contains a listing of all facilities in a given State with a list of the chemicals they reported. States can use this report as an initial screening tool to ensure that their list of submissions matches EPA's. The reports show the following information:

- State "Top 100" On-site Report. The State "Top 100" On-site Report is a listing of the 100 largest on-site emitting facilities in a given State.
- State "Top 100" On- and Off-site Report. The State "Top 100" On- and Off-site Report is a listing of the 100 largest emitting facilities (combining both on- and off-site emissions) in a given State.
- State Chemical Report. The State Chemical Report will show chemicals ranked by total releases in a given state.

Through the State Reconciliation Exercise in the Fall of each year, EPA learns from States about submissions the State may have that EPA does not have and vice versa. When EPA learns of submissions received by the State and not by EPA, a "Failure to Submit to EPA" letter is mailed to the facility. Facilities need to report to both EPA and the State. In addition, EPA may learn of any discrepancies in data reported to the State versus EPA.

*EPA would like comments on all aspects of the State Reconciliation Report process. Suggestions*

*and comments can discuss the effectiveness of this report to inform EPA and assist States.*

## **Step 6: Data “Chill” and Data Quality Steps**

This step is intended to identify and request the correction of significant submitter errors in their releases and other waste management quantities.

Prior to freezing the TRI data for release to the public, EPA (both headquarters and regions) reviews the data for data quality in a number of ways. Certain data quality activities are done each year, while others are dependent on the particular reporting year. The first step for data quality purposes is the delivery of a “chill” version of TRIS to EPA. The chill is the first snapshot of all data received and entered, together with as many FDP responses received and entered in TRIS.

*Significant Increases/Decreases and Largest Releases and Production-related Waste:* EPA focuses significant effort on reviewing reports submitted by facilities reporting large increases and decreases in releases and production-related waste (both on- and off-site releases, and total production-related waste) as compared to the previous reporting year. EPA also focuses on facilities reporting the largest total releases and production-related waste. By ranking facilities in largest ranked industry sectors (i.e., ranked by total releases and total production-related waste), EPA has been able to discover large errors that significantly skew both trend and single year analyses, optimizing limited time and resources.

*Other Data Quality:* Depending upon the reporting year, EPA also customizes data quality efforts in areas that may be prone to significant errors. For example, in preparation for the 2000 TRI data release, EPA focused significant attention on several TRI industry sector categories, including: sectors reporting significant errors in the past, sectors recently required to report to TRI (i.e., facility expansion in 1998), and sectors reporting large quantities of persistent, bioaccumulative toxic (PBT) chemicals (PBT chemicals added/lowered thresholds for the 2000 reporting year). By focusing on such reports, EPA has been able to optimize data quality efforts with available resources.

For example, in preparation for the 2000 data release, EPA headquarters and regions called more than 200 facilities that met the largest increases/decreases and largest release/waste criteria. Approximately 27 of the facilities contacted by EPA reported at least one significant error that needed to be corrected. Due to the reporting of PBT chemicals at lowered thresholds in 2000, EPA also focused its efforts on PBTs chemicals and contacted 560 facilities reporting such chemicals. As a result, approximately 130 facilities revised their reported release and other waste management data for PBT chemicals. In another example, facilities revised their total release quantity of dioxin and dioxin-like compounds to 95,910 grams from the 750,226 grams originally reported. In addition, EPA TRI regional programs also carried out various data quality analyses in preparation for the 2000 data release.



*EPA would like comments regarding “Data Chill” and its Data Quality Procedures.*

### **Step 7: Data “Freeze”**

This step is intended halt any new information from being entered into the TRIS database. This allows the Reporting Center to prepare the TRI data for the Public Data Release.

After the data “chill” and data quality activities and corrections have been entered, the TRIS database is then “frozen” in order to prepare the data for the TRI public data release. Once the TRI data are Afrozen,@it takes the Agency approximately 3 months to conduct the analyses, prepare the data release reports, and prepare the data access tools to be ready to Arelease@the data to the public. For more information in regards to the Data “Freeze” and post “freeze” activities, refer to the issue paper *TRI Data Release Issue Paper*.

## **IV. REQUEST FOR COMMENTS**

We would like to hear from you if you are aware of other ways to process data, rather than the established EPA Process. At a minimum, the modeled process should meet the following important aspects for TRI data processing:

- Process 120,000 documents and forms
- Improve data quality
- Ensure and promote reporter accountability
- Allow electronic corrections to forms to expedite corrections by submitters
- Internet reporting capabilities
- Accelerate Public Data Release

# TRIS Data Flow

